
Book of Abstracts and Posters



Metadata and
Semantics Research
12th International Research Conference,
MTSR 2018
October 23rd-26th, 2018
Limassol, Cyprus

Book of Abstracts

Editing and Design

Emmanouel Garoufallou

Working Group

Iro Sotiriadou, Anxhela Dani, Marios Zervas, Rania Siatri, Fabio Sartori, Damiana Koutsomiha, Stavroula Antonopoulou, Chrysanthi Chatzopoulou

Co-organized by

Library and Information
Services | Cyprus University
of Technology



Cyprus
University of
Technology

Library and
Information
Services

Department of Informatics,
Systems and
Communication | University
of Milano-Bicocca, Italy



School of Information |
Kent State University



Marconi University



Georg Eckert Institute



Department of Library
Science and Information
Systems | Alexander TEI of
Thessaloniki



Department of Library Science
and Information Systems
Alexander TEI of Thessaloniki

Alexander Technological
Educational Institute of
Thessaloniki



Awards sponsored by

euroCRIS
Current Research
Information Systems



Sponsors

Visit Cyprus



Passion for Success | Cyprus
Convention Bureau



Elsevier



Springer Nature



EBSCO



ACS Publications



IEEE



ReasonableGraph



Cataloging-in-Publication Data

International Conference on Metadata and Semantics Research (12th : 2018 : Limassol, Cyprus)

Metadata and semantics research : 12th Research Conference, MTSR 2018, October 23rd-26th October 2018, Limassol, Cyprus : book of abstracts / editing and design Emmanouel Garoufallou ; working group Iro Sotiriadou ... [et al.]

p. cm.

Includes index.

ISBN: 978-9963-697-34-2

1. Metadata—Congresses. 2. Semantic web—Congresses. 3. Semantics—Data processing—Congresses. 4. Semantic computing—Congresses. I. Garoufallou, Emmanouel. II. Sotiriadou, Iro.

006'.7'—dc23

Table of Contents

<i>Table of Contents</i>	7
Preface	9
Organization	12
Keynote Speaker	23
Abstracts	25
General Session	26
<i>A Semantic Web SKOS Vocabulary Service for Open Knowledge Organization Systems</i>	26
<i>Document Based RDF Storage Method for Efficient Parallel Query Processing</i>	26
<i>Legal Entity Identifier Blockchained by a Hyperledger Indy Implementation of</i> <i>GraphChain</i>	27
<i>Query Translation for Cross-lingual Search in the Academic Search Engine PubPsych</i>	27
<i>ViziQuer: a visual notation for RDF data analysis queries</i>	28
<i>A Systematic Approach to Review Legacy Schemas Based on Ontological Analysis</i>	28
<i>Towards a holistic documentation and wider use of Digital Cultural Heritage</i>	29
<i>Graph Matching based Semantic Search Engine</i>	30
<i>SKOS-based Concept Expansion for LOD-enabled Recommender Systems</i>	30
<i>Navigating OWL 2 Ontologies through Graph Projection*</i>	31
<i>Linked Data Live Exploration with Complete Results</i>	31
Track on Digital Libraries, Information Retrieval, Big, Linked, Social and Open Data	34
<i>Analysing and Visualising Open Data within the Data & Analytics Framework</i>	34
<i>Formalizing Enrichment Mechanisms for Bibliographic Ontologies in the Semantic</i> <i>Web</i>	34
<i>GLOBDEF: a framework for dynamic pipelines of semantic data enrichment tools</i>	35
<i>Ontologies for Data Science: on its application to data pipelines</i>	36
<i>Relating Legal Entities via Open Information Extraction</i>	36
<i>Ontology-based Information Retrieval: development of a semantic-based tool for the</i> <i>media industry</i>	37
Track on Cultural Collections and Applications	38
<i>Evaluating Data Quality in Europeana: metrics for multilinguality</i>	38
<i>The Benefits of Linking Metadata for Internal and External Users of an Audiovisual</i> <i>Archive</i>	38
<i>Authify: the reconciliation of entities at scale</i>	39

<i>Assessing the Preservation of Derivative Relationships in Mappings from FRBR to BIBFRAME</i>	<i>40</i>
<i>Metadata Standards for Palm Leaf Manuscripts in Asia</i>	<i>40</i>
Track on Knowledge IT Artifacts (KITA) in Professional Communities and Aggregations (KITA 2018)	41
<i>Knowledge Artifacts for the Health: the PERCIVAL project</i>	<i>41</i>
<i>Artfacts: a platform for making sense of and telling stories with cultural objects</i>	<i>41</i>
<i>A Semantic-based Metadata Schema to Handle System and Software Configuration.</i>	<i>42</i>
Track on Digital Humanities and Digital Curation (DHC)	43
<i>Connecting and Mapping LOD and CMDI through Knowledge Organization</i>	<i>43</i>
<i>Creating CMDI-Profiles for Textbook Resources.....</i>	<i>43</i>
Track on European and National Projects	45
<i>Towards a Knowledge Graph based Platform for Public Procurement*</i>	<i>45</i>
<i>Metadata for Large-scale Research Instruments</i>	<i>45</i>
Track on Agriculture, Food and Environment	47
<i>Identification and Exchange of Regulated Information on Chemicals: from metadata to Core Vocabulary*</i>	<i>47</i>
<i>Semantics for Data in Agriculture: a community-based wish list</i>	<i>47</i>
<i>Development of Methodologies and Standardized Services for Supporting Forest Economies</i>	<i>48</i>
Track on Open Repositories, Research Information Systems and Data Infrastructures	50
<i>Open Citation Content Data</i>	<i>50</i>
<i>The Case for Ontologies in Expressing Decisions in Decentralized Energy Systems</i>	<i>50</i>
Posters	52
<i>Metadata Based Datasets Placement in Smart Grids.....</i>	<i>53</i>
<i>EFSA Controlled Terminology – FoodEx2 Classification and Description System</i>	<i>57</i>
<i>Set Up of a Metadata Strategy – First step: indexing your data</i>	<i>61</i>
Author Index.....	63

Preface

Since 2005, the International Metadata and Semantics Research Conference (MTSR) has served as a significant venue for dissemination and sharing of metadata and semantic-driven research and practices. This year, 2018, marks the 12th MTSR, drawing scholars, researchers and practitioners who are investigating and advancing our knowledge on a wide range of metadata and semantic-driven topics. The 12th International Conference on Metadata and Semantics Research (MTSR'18) was held at the Cyprus University of Technology in Cyprus, between 23-26 October 2018.

Metadata and semantics are integral to any information system and important to the sphere of Web data. Research and development addressing metadata and semantics is crucial to advancing how we effectively discover, use, archive, and repurpose information. In response to this need, researchers are actively examining methods for generating, reusing, and interchanging metadata. Integrated with these developments is research on the application of computational methods, linked data, and data analytics. A growing body of literature also targets conceptual and theoretical designs providing foundational frameworks for metadata and semantic applications. There is no doubt that metadata weaves its way through nearly every aspect of our information ecosystem, and there is great motivation for advancing the current state of understanding in the fields of metadata and semantics. To this end, it is vital that scholars and practitioners convene and share their work.

MTSR conferences have grown in number of participants and paper submission rates over the last decade, marking it as a leading, international research conference. Continuing in the successful legacy of previous MTSR conferences (MTSR 2005, MTSR 2007, MTSR 2009, MTSR 2010, MTSR 2011, MTSR 2012, MTSR 2013, MTSR 2014, MTSR 2015, MTSR 2016 and MTSR 2017), MTSR 2018 brings together scholars and practitioners who share a common interest in the interdisciplinary field of metadata, linked data, and ontologies.

The MTSR 2018 programme and the following proceedings show a rich diversity of research and practices from metadata and semantically focused tools and technologies, linked data, cross language semantics, ontologies, metadata models, semantic systems, and metadata standards. The general session of the conference included 12 papers covering a broad spectrum of topics, proving the interdisciplinary view of metadata. Metadata as a research topic is maturing, and the conference supported the following seven tracks: Digital Libraries, Information Retrieval, Big, Linked, Social & Open Data; Metadata and Semantics for Cultural Collections and Applications; Metadata and Semantics for Open Repositories, Research Information Systems and Data Infrastructures; Track on Digital Humanities and Digital Curation; Metadata and Semantics for Agriculture, Food, and Environment; Track on Knowledge IT Artifacts in professional communities and aggregations and Track on European and National Projects. Each of these tracks had a rich selection of short and full research papers, in total 23, giving broader diversity to MTSR, and enabling deeper exploration of significant topics.

All the papers underwent a thorough and rigorous peer-review process. The review and selection for this year was highly competitive and only papers containing significant research results, innovative methods, or novel and best practices were accepted for publication. From the general session, only 8 submissions were accepted as full research papers representing 26,6% of the total number of submissions and 4 as short papers. Additional 11 contributions from tracks covering noteworthy and important results were accepted as full research papers representing 23,4% of the total number of submissions, and 12 as short papers, totaling for this year's MTSR 2018 accepted contributions to 35. The acceptance rate of full research papers for both the general session and tracks was 24,6% of the total number of submissions, while the cumulative percentage of acceptance rate for all papers was 45,5%.

Cyprus University of Technology aspires to become a modern and innovative university with international recognition which promotes excellence in education and research in cutting-edge fields, aiming at the scientific, technological, economic, social and cultural upgrading of our country. Our mission is to provide high-quality education and high-level training and to promote lifelong learning with modern pedagogical methods, to produce and disseminate scientific knowledge through research and teaching, and through contribution towards addressing challenges in the fields of Science and Technology, to become a catalyst for the state and society. In particular, the mission focuses on creating added value through actions on the following three axes: Education, Research, Innovation and Transfer of Know-how and Social contribution. The Library's mission is to serve and support the members of the academic society, enhancing the scientific research and the educational procedure. The Library actively participates to the University's mission for providing education and research of high level in important sectors. This fact is achieved with the access to information and to the registered knowledge, with the carrying out of seminars as well as with the ensuring and organization of noteworthy source of information for the development of research and other skills from its users-members.

This year the MTSR conference was pleased to host one remarkable keynote presentation by Dr Christos Papatheodorou, Professor in the Department of Archives, Library Science and Museology at the Ionian University in Greece. In his presentation "Library Data models under the lens of interoperability and quality", Professor Papatheodorou shared his extensive experience and insights about metadata quality approaches and their adaptation to the various models of bibliographic data and their interoperability.

We conclude this preface by thanking the many people who contributed their time and efforts to MTSR 2018 and made this year's conference possible. We also thank all the organizations that supported this conference. We extend a sincere gratitude to members of the Programme Committees both main and special tracks, the Steering Committee and the Organizing Committees (both general and local), and the conference reviewers who invested their time generously to ensure the timely review of the submitted manuscripts. A special thank to Programme Chairs Ernesto William De Luca from Georg Eckert-Institute – Leibniz-Institute for International Textbook Research, Germany and Athena Salaba from Kent State University, USA; to Workshop, Tutorial and Demonstration Chair Dr Getaneh Alemu from Southampton

Solent University, UK; to Anxhela Dani and Marina Antoniou for supporting us throughout this event and to Iro Sotiriadou and Anxhela Dani who assisted us with the preparation of proceedings; and to Nikoleta, Vasiliki and Stavroula for their endless support and patience. Our thanks go to SPRINGER NATURE, ELSEVIER, EBSCO, ACS, IEEE, Cyprus Tourism Organisation and Reasonable Graph. Final, our thanks go to all participants of MTSR 2018 for making the event a great success.

November 2018

Emmanouel Garoufallou

Fabio Sartori

Rania Siatiri

Marios Zervas

Organization

General Chairs

Emmanouel Garoufallou	Alexander Technological Educational Institute (ATEI) of Thessaloniki, Greece
Fabio Sartori	University of Milano-Bicocca, Italy

Program Chairs

Ernesto William De Luca	Georg Eckert-Institute – Leibniz-Institute for International Textbook Research, Germany
Athena Salaba	Kent State University, USA

Local Organization Chair

Marios Zervas	Cyprus University of Technology, Cyprus
---------------	---

Workshop, Tutorial and Demonstration Chair

Getaneh Alemu	Southampton Solent University, UK
---------------	-----------------------------------

Special Track Chairs

Paolo Bianchini	Università degli Studi di Torino, Italy
Miguel-Ángel Sicilia	University of Alcalá, Spain
Armando Stellato	University of Rome Tor Vergata, Italy
Angela Locoro	University of Milano-Bicocca, Italy
Arlindo Flavio da Conceição	Federal University of São Paulo (UNIFESP), Brazil
Nikos Houssos	IRI, Greece
Michalis Sfakakis	Dept. Archives, Library Science and Museology, Ionian University, Corfu, Greece
Lina Bountouri	Dept. Archives, Library Science and Museology, Ionian University, Corfu, Greece
Emmanouel Garoufallou	Alexander Technological Educational Institute (ATEI) of Thessaloniki, Greece
Sirje Virkus	Tallinn University, Estonia

Rania Siatri	Alexander Technological Educational Institute (ATEI) of Thessaloniki, Greece
Ernesto William De Luca	Georg Eckert-Institute – Leibniz-Institute for International Textbook Research, Germany
Fabio Sartori	University of Milano-Bicocca, Italy

Steering Committee

Juan Manuel Dodero	University of Cádiz, Spain
Emmanouel Garoufallou	Alexander Technological Educational Institute (ATEI) of Thessaloniki, Greece
Nikos Manouselis	AgroKnow, Greece
Fabio Santori	University of Milano-Bicocca, Italy
Miguel-Ángel Sicilia	University of Alcalá, Spain

Organizing Committee

Pavlos Metaxas	Cyprus University of Technology, Cyprus
Stephani Liasi	Cyprus University of Technology, Cyprus
Maria Haraki	Cyprus University of Technology, Cyprus
Stamatios Giannoulakis	Cyprus University of Technology, Cyprus
Eleni Kamberi	Cyprus University of Technology, Cyprus
Nasia Panagiotou	Cyprus University of Technology, Cyprus
Marina Antoniou	Cyprus University of Technology, Cyprus
Alexia Kounoudes	Cyprus University of Technology, Cyprus
Antonia Nikolaou	Cyprus University of Technology, Cyprus
Damiana Koutsomiha	American Farm School, Greece
Anxhela Dani	Alexander Technological Educational Institute (ATEI) of Thessaloniki, Greece
Chrysanthi Chatzopoulou	Alexander Technological Educational Institute (ATEI) of Thessaloniki, Greece
Iro Sotiriadou	American Farm School, Greece

Technical Support Staff

Ilias Nitsos	Alexander Technological Educational Institute (ATEI) of Thessaloniki, Greece
--------------	--

Petros Artemi

Cyprus University of Technology, Cyprus

Program Committee

Rajendra Akerkar

Western Norway Research Institute, Norway

Arif Altun

Hacettepe University, Turkey

Ioannis N. Athanasiadis

Democritus University of Thrace, Greece

Panos Balatsoukas

University of Manchester, UK

Tomaz Bartol

University of Ljubljana, Slovenia

Ina Bluemel

German National Library of Science and
Technology TIBm, Germany

Derek Bousfield

Manchester Metropolitan University, UK

Gerhard Budin

University of Vienna, Austria

Özgü Can

Ege University, Turkey

Caterina Caracciolo

Food and Agriculture Organization (FAO) of the
United Nations, Italy

Christian Cechinel

Federal University of Pampa, Brazil

Artem Chebotko

University of Texas - Pan American, USA

Philip Cimiano

Bielefeld University, Germany

Sissi Closs

Karlsruhe University of Applied Sciences,
Germany

Ricardo Colomo-Palacios

Universidad Carlos III, Spain

Sally Jo Cunningham

Waikato University, New Zealand

Constantina Costopoulou

Agricultural University of Athens, Greece

Ernesto William De Luca

Georg Eckert-Institute – Leibniz-Institute for
International Textbook Research, Germany

Milena Dobрева

University of Malta, Malta

Juan Manuel Dodero

University of Cádiz, Spain

Erdogan Dogdu

Cankaya University, Turkey

Biswanath Dutta

Documentation Research and Training Centre
(DRTC), Indian Statistical Institute, India

Juan José Escribano Otero

Universidad Europea de Madrid, Spain

Muriel Foulonneau

Tudor Public Research Centre, Luxemburg

Panorea Gaitanou

Ionian University, Greece

Emmanouel Garoufalou

Alexander Technological Educational Institute
(ATEI) of Thessaloniki, Greece

Manolis Gergatsoulis

Ionian University, Greece

Jorge Gracia	University of Zaragoza, Spain
Jane Greenberg	Drexel University, USA
Jill Griffiths	Manchester Metropolitan University, UK
R. J. Hartley	Manchester Metropolitan University, UK
Nikos Houssos	IRI, Greece
Carlos A. Iglesias	Universidad Politécnica de Madrid, Spain
Frances Johnson	Manchester Metropolitan University, UK
Dimitris Kanellopoulos	University of Patras, Greece
Pinar Karagöz	Middle East Technical University (METU), Turkey
Pythagoras Karampiperis	AgroKnow, Greece
Brian Kelly	CETIS, University of Bolton, UK
Christian Kop	University of Klagenfurt, Austria
Rebecca Koskela	University of New Mexico, USA
Daniela Luzi	National Research Council, Italy
Paolo Manghi	Institute of Information Science and Technologies (ISTI), National Research Council (CNR), Italy
John McCrae	National University of Ireland Galway, Ireland
Xavier Ochoa	Centro de Tecnologías de Información Guayaquil, Ecuador
Mehmet C. Okur	Yaşar University, Turkey
Matteo Palmonari	University of Milano-Bicocca, Italy
Manuel Palomo Duarte	University of Cádiz, Spain
Laura Papaleo	University of Genova, Italy
Christos Papatheodorou	Ionian University, Greece
Marios Poulos	Ionian University, Greece
T. V. Prabhakar	Indian Institute of Technology Kanpur, India
Maria Cláudia Reis Cavalcanti	Military Institute of Engineering, Brazil
Athena Salaba	Kent State University, USA
Salvador Sánchez-Alonso	University of Alcalá, Spain
Fabio Sartori	University of Milano-Bicocca, Italy
Cleo Sgouropoulou	Technological Educational Institute of Athens, Greece
Rania Siatri	Alexander Technological Educational Institute (ATEI) of Thessaloniki, Greece
Miguel-Ángel Sicilia	University of Alcalá, Spain

Armando Stellato	University of Rome Tor Vergata, Italy
Imma Subirats	Food and Agriculture Organization (FAO) of the United Nations, Italy
Shigeo Sugimoto	University of Tsukuba, Japan
Stefaan Ternier	Open University of the Netherlands, Netherlands
Giannis Tsakonas	University of Patras, Greece
Andrea Turbati	University of Rome Tor Vergata, Italy
Fabio Massimo Zanzotto	University of Rome Tor Vergata, Italy
Thomas Zschocke	World Agroforestry Centre (ICRAF), Kenya

Track on Metadata and Semantics for Digital Libraries, Information Retrieval, Big, Linked, Social and Open Data

Special Track Chairs

Emmanouel Garoufallou	Alexander Technological Educational Institute (ATEI) of Thessaloniki, Greece
Rania Siatiri	Alexander Technological Educational Institute (ATEI) of Thessaloniki, Greece
Sirje Virkus	Tallinn University, Estonia

Program Committee

Panos Balatsoukas	City University, UK
Özgü Can	Ege University, Turkey
Sissi Closs	Karlsruhe University of Applied Sciences, Germany
Mike Conway	University of North Carolina at Chapel Hill, USA
Phil Couch	University of Manchester, UK
Milena Dobрева	University of Malta, Malta
Ali Emrouznejad	Aston University, UK
Panorea Gaitanou	Ionian University, Greece
Stamatios Giannoulakis	Cyprus University of Technology, Cyprus
Jane Greenberg	Drexel University, USA
R. J. Hartley	Manchester Metropolitan University, UK
Nikos Korfiatis	University of East Anglia, UK
Rebecca Koskela	University of New Mexico, USA
Valentini Moniarou-Papaconstantinou	Technological Educational Institute of Athens, Greece
Dimitris Rousidis	University of Alcalá, Spain
Christine Urquhart	Aberystwyth University, UK
Evgenia Vassilakaki	Technological Educational Institute of Athens, Greece
Sirje Virkus	Tallinn University, Estonia
Georgia Zafeiriou	University of Macedonia, Greece

Track on Metadata and Semantics for Cultural Collections and Applications

Special Track Chairs

Michalis Sfakakis	Dept. Archives, Library Science and Museology, Ionian University, Corfu, Greece
Lina Bountouri	Dept. Archives, Library Science and Museology, Ionian University, Corfu, Greece

Program Committee

Trond Aalberg	Norwegian University of Science and Technology (NTNU), Norway
Karin Bredenberg	The National Archives of Sweden, Sweden
Enrico Francesconi	EU Publications Office, Luxembourg, and Consiglio Nazionale delle Ricerche, Firenze, Italy
Manolis Gergatsoulis	Ionian University, Greece
Antoine Isaac	Vrije Universiteit, Amsterdam, the Netherlands
Sarantos Kapidakis	Ionian University, Greece
Peter McKinney	National Library of New Zealand Te Puna Mātauranga o Aotearoa, New Zealand
Christos Papatheodorou	Ionian University and Digital Curation Unit, IMIS, Athena RC, Greece
Chrisa Tsinaraki	Joint Research Centre, European Commission, Italy
Andreas Vlachidis	Faculty of Computing, Engineering and Science, University of South Wales, UK
Katherine Wisser	Graduate School of Library and Information Science, Simmons College, USA
Maja Žumer	University of Ljubljana, Slovenia

Track on Metadata and Semantics for Knowledge IT Artifacts (KITA) in Professional Communities and Aggregations (KITA 2018)

Special Track Chairs

Fabio Sartori	University of Milano-Bicocca, Italy
Angela Locoro	University of Milano-Bicocca, Italy
Arlindo Flavio da Conceição	Federal University of São Paulo (UNIFESP), Brazil

Program Committee

Federico Cabitza	University of Milano-Bicocca, Italy
Luca Grazioli	ICteam SpA, Italy
Riccardo Melen	University of Milano-Bicocca, Italy
Aurelio Ravarini	Università Carlo Cattaneo – LIUC, Castellanza, Italy
Carla Simone	University of Siegen, Germany
Flávio Soares Corrêa da Silva	University of São Paulo, Brazil
Cecilia Zanni-Merk	Insa Rouen Normandie, France

Track on Metadata and Semantics for Digital Humanities and Digital Curation (DHC)

Special Track Chairs

Ernesto William De Luca	Georg Eckert Institute – Leibniz Institute for International Textbook Research, Germany
Paolo Bianchini	Università degli Studi di Torino, Italy

Program Committee

Maret Keller	Georg Eckert Institute – Leibniz Institute for International Textbook Research, Germany
Elena González-Blanco	Universidad Nacional de Educación a Distancia, Spain
Steffen Hennicke	Georg Eckert Institute – Leibniz Institute for International Textbook Research, Germany
Ana García-Serrano	ETSI Informatica – UNED, Spain

Andreas Weiß	Georg Eckert Institute – Leibniz Institute for International Textbook Research, Germany
Philipp Mayr	GESIS, Germany
Lena-Luise Stahn	Georg Eckert Institute – Leibniz Institute for International Textbook Research, Germany
Andrea Turbati	University of Rome Tor Vergata, Italy
Christian Scheel	Georg Eckert Institute – Leibniz Institute for International Textbook Research, Germany
Armando Stellato	University of Rome Tor Vergata, Italy
Wolf-Tilo Balke	TU Braunschweig, Germany
Andreas Lommatzsch	TU Berlin, Germany
Ivo Keller	TH Brandenburg, Germany
Gabriela Ossenbach	UNED, Spain
Francesca Fallucchi	Guglielmo Marconi University, Italy
Alessandra Pieroni	Guglielmo Marconi University, Italy

Track on Metadata and Semantics for European and National Projects

Special Track Chairs

Emmanouel Garoufallou	Alexander Technological Educational Institute (ATEI) of Thessaloniki, Greece
Stavroula Antonopoulou	American Farm School, Greece

Program Committee

Panos Balatsoukas	City University, UK
Mike Conway	University of North Carolina at Chapel Hill, USA
Jane Greenberg	Drexel University, USA
R. J. Hartley	Manchester Metropolitan University, UK
Nikos Houssos	IRI, Greece
Nikos Korfiatis	University of East Anglia, UK
Damiana Koutsomiha	American Farm School, Greece
Paolo Manghi	Institute of Information Science and Technologies (ISTI), National Research Council (CNR), Italy
Dimitris Rousidis	University of Alcalá, Spain

Rania Siatra	Alexander Technological Educational Institute (ATEI) of Thessaloniki, Greece
Miguel-Ángel Sicilia	University of Alcalá, Spain
Armando Stellato	University of Rome Tor Vergata, Italy
Sirje Virkus	Tallinn University, Estonia

Track on Metadata and Semantics for Agriculture, Food and Environment (AgroSEM'18)

Special Track Chairs

Miguel-Ángel Sicilia	University of Alcalá, Spain
----------------------	-----------------------------

Program Committee

Ioannis Athanasiadis	Wageningen University, the Netherlands
Patrice Buche	INRA, France
Caterina Caracciolo	Food and Agriculture Organization (FAO) of the United Nations, Italy
Johannes Keizer	Food and Agriculture Organization (FAO) of the United Nations, Italy
Stasinos Konstantopoulos	NCSR "Demokritos", Greece
Claire Nédellec	INRA, France
Armando Stellato	University of Rome Tor Vergata, Italy
Maguelonne Teisseire	Irstea Montpellier, France
Jan Top	Wageningen Food & Biobased Research, the Netherlands
Robert Trypuz	John Paul II Catholic University of Lublin, Poland

Track on Metadata and Semantics for Open Repositories, Research Information Systems and Data Infrastructures

Special Track Chairs

Nikos Houssos	IRI, Greece
Armando Stellato	University of Rome Tor Vergata, Italy

Honorary Track Chairs

Imma Subirats

Food and Agriculture Organization (FAO) of the
United Nations, Italy

Program Committee

Sophie Aubin

INRA, France

Thomas Baker

Sungkyunkwan University, Korea

Hugo Besemer

Wageningen UR Library, the Netherlands

Gordon Dunshire

University of Strathclyde, UK

Jan Dvorak

Charles University of Prague, Czech Republic

Jane Greenberg

Drexel University, USA

Siddeswara Guru

University of Queensland, Australia

Keith Jeffery

Keith G. Jeffery Consultants, UK

Nikolaos Konstantinou

University of Manchester, UK

Rebecca Koskela

University of New Mexico, USA

Jessica Lindholm

Malmö University, Sweden

Paolo Manghi

Institute of Information Science and
Technologies - Italian National Research
Council (ISTI-CNR), Italy

Brian Matthews

Science and Technology Facilities Council, UK

Eva Mendez Rodriguez

Universidad Carlos III of Madrid, Spain

Joachim Schöpfel

University of Lille, France

Kathleen Shearer

Confederation of Open Access Repositories
(COAR), Germany

Jochen Schirrwagen

University of Bielefeld, Germany

Birgit Schmidt

University of Göttingen, Germany

Chrisa Tsinaraki

European Commission, Joint Research Centre,
Italy

Yannis Tzitzikas

University of Crete and ICS-FORTH, Greece

Zhong Wang

Sun-Yat-Sen University, China

Marcia Zeng

Kent State University, USA

Keynote Speaker



Prof. Christos Papatheodorou

Department of Archives, Library Science and Museology - Ionian University

Christos Papatheodorou is a professor of Information Systems for Libraries and Archives at the Department of Archives, Library Science and Museology, Ionian University, Greece and co-leader of the Database and Information Systems Group (DBIS), Digital Libraries and Electronic Publishing Laboratory of the same Department. He is also fellow researcher in the Digital Curation Department, Institute for the Management of Information Systems, "Athena" Research Centre, Athens, Greece. He holds a B.Sc. and a Ph.D. in Computer Science. His research interests include Evaluation of Digital Libraries and Information Services, Metadata interoperability, Web 2.0 technologies, Digital Preservation, Personalized access to Digital Libraries and Internet, User Modeling and Web Mining. He is Associate Editor of the International Journal of Metadata, Semantics and Ontologies and he was the Chair of the Steering Committee of the International Conference on Theory and Practice of Digital Libraries (TPDL), from September 2012 to September 2015.

Keynote title: Library Data models under the lens of interoperability and quality

Any type of digital information that describes resources or supports their discovery and is produced or curated by libraries is defined as Library Data. Indicatively such information could be datasets (e.g. catalogues, bibliographies), vocabularies and metadata elements (e.g. elements of MODS or MADS). Nowadays these data are published as Linked Data and thus are called Library Linked Data (LLD). LLD exploit the benefits of the Web technologies and provide to their users improved capabilities for discovering library and non-library information resources from other trusted sources. Moreover, the cataloguing paradigm changes from top-down, where libraries produce records for their material, to bottom-up, where resources can be described in collaboration with libraries and also linked to data contributed by other communities.

There is a number of initiatives regarding the exposure of LLD into the Web. Depending on the initiative's objectives, the scope and the intended use of the data, each initiative

developed its own interpretation of how Library Data may be integrated into the semantic web, providing its own conceptual model, such as FRBR, BIBFRAME and EDM. However, these different views cause interoperability problems and prevent data integration.

Moving towards a data driven world, the quality of data is critical and affects the outcome of all data management related tasks. Metadata quality also affects information integration and retrieval since they are metadata driven, while its evaluation should consider their interpretation as determined by the conceptual model in use. This presentation aims to interlink the concepts of LLD interoperability and data quality. It reveals LLD interoperability issues providing cases of bibliographic patterns that are represented differently by particular library data conceptual models. Furthermore, it overviews results from ongoing research regarding the assurance of the quality of Library Data during their integration, when instances of one library conceptual model are transformed to those of another.

Abstracts

General Session

A Semantic Web SKOS Vocabulary Service for Open Knowledge Organization Systems

Jonas Waeber and Andreas Ledl^[0000-0002-0629-0446]

Basel University Library, 4056 Basel, Switzerland
andreas.ledl@unibas.ch

Abstract: In this article, the Basel Register of Thesauri, Ontologies & Classifications (BARTOC.org) is introduced to raise awareness for an integrated, full terminology registry for knowledge organization systems. Recently, researchers have shown an increased interest in such a single access point for controlled vocabularies. The paper outlines BARTOC's technical implementation, system architecture, and services in the light of semantic technologies. Its central thesis is that if the KOS community agreed on BARTOC as one of their main terminology registries, all involved parties would benefit from linked open knowledge organization systems.

Keywords: Knowledge organization systems, terminology registry, BARTOC, Skosmos, SKOS, semantic web.

Document Based RDF Storage Method for Efficient Parallel Query Processing

Eleftherios Kalogeros, Manolis Gergatsoulis, and Matthew Damigos

Database & Information Systems Group (DBIS),
Department of Archives, Library Science and Museology, Ionian University, Greece
{kalogero, manolis}@ionio.gr, mgdamig@gmail.com

Abstract: In this paper, we investigate the problem of efficiently evaluating SPARQL queries, over large amount of linked data utilizing distributed NoSQL system. We propose an efficient approach for partitioning large linked data graphs using distributed frameworks (MapReduce), as well as an effective data model for storing linked data in a document database using a maximum replication factor of 2 (i.e., in the worst case scenario, the data graph will be doubled in storage size). The model proposed and the partitioning approach ensure high-performance query evaluation and horizontal scaling for the type of queries called *generalized star queries* (i.e., queries allowing both subject-object and object-subject edges from a central node), due to the fact that no joining operations over multiple datasets are required to evaluate the queries. Furthermore, we present an implementation of our approach using

MongoDB and an algorithm for translating generalized star queries into MongoDB query language, based on the proposed data model.

Keywords: RDF, linked data, parallel processing, NoSQL, document databases.

Legal Entity Identifier Blockchained by a Hyperledger Indy Implementation of GraphChain

Mirek Sopek¹, Przemysław Grądzki^{1,2}, Dominik Kuziński¹, Rafał Trójczak^{1,2}, and Robert Trypuz^{1,2}

¹ R&D Team, MakoLab S.A.,
ul. Demokratyczna 46,
93-430 Łódź, Poland
sopek@makolab.com

² The John Paul II Catholic University of Lublin,
Faculty of Philosophy,
Al. Racławickie 14,
20-950 Lublin, Poland
robert.trypuz@makolab.com

Abstract: The main idea behind GraphChain is to use blockchain mechanisms on top of abstract RDF graphs. This paper presents an implementation of GraphChain in the Hyperledger Indy framework. The whole setting is shown to be applied to the RDF graphs containing information about Legal Entity Identifiers (LEIs). The blockchain based data management system presented in the paper preserves all the benefits of using RDF data model for the representation of LEI system reference data, including powerful querying mechanisms, explicit semantics and data model extensibility with the security and non-repudiation of LEIs as the digital identifiers for legal entities.

Keywords: Hyperledger, Hyperledger Indy, GraphChain, semantic blockchain, LEI, GLEIS.

Query Translation for Cross-lingual Search in the Academic Search Engine PubPsych

Cristina España-Bonet¹ [0000-0001-5414-4710], Juliane Stiller² [0000-0001-8184-6187],
Roland Ramthun³ [0000-0003-3300-6575], Josef van Genabith¹, and Vivien Petras² [0000-0002-8113-1509]

¹ Universität des Saarlandes & DFKI, Saarbrücken, Germany
cristinae@dfki.de

² Berlin School of Library and Information Science,
Humboldt-Universität zu Berlin, Germany

Abstract: We describe a lexical resource-based process for query translation of a domain-specific and multilingual academic search engine in psychology, PubPsych. PubPsych queries are diverse in language with a high amount of informational queries and technical terminology. We present an approach for translating queries into English, German, French, and Spanish. We build a quadrilingual lexicon with aligned terms in the four languages using MeSH, Wikipedia and Apertium as our main resources. Our results show that using the quadlexicon together with some simple translation rules, we can automatically translate 85% of translatable tokens in PubPsych queries with mean adequacy over all the translatable text of 1.4 when measured on a 3-point scale [0,1,2].

Keywords: Academic search, psychology domain, logfile analysis, query languages, query translation, translation quality, digital library.

ViziQuer: a visual notation for RDF data analysis queries

Kārlis Čerāns^{1,2,4}, Agris Šostaks^{1,2}, Uldis Bojārs^{1,2}, Juris Bārzdīņš³, Jūlija Ovčinnikova^{1,2},
Mikus Grasmanis¹, Artūrs Sproģis¹

¹ Institute of Mathematics and Computer Science, University of Latvia

² Department of Computing, University of Latvia

³ Department of Medicine, University of Latvia

⁴ karlis.cerans@lumii.lv

Abstract: Visual SPARQL query notations aim at easing the RDF data querying task. At the current state of the art there is still no generally accepted visual graph-based notation suitable to describe RDF data analysis queries that involve aggregation and subqueries. In this paper we present a visual diagram-centered notation for SPARQL select query formulation, capable to handle aggregate/statistics queries and hierarchic queries with subquery structure. The notation is supported by a web-based prototype tool. We present the notation examples, describe its syntax and semantics and describe studies with possible end users, involving both IT and medicine students.

Keywords: Visual notation, diagrammatic queries, RDF data, SPARQL, ad-hoc queries, data analysis queries.

A Systematic Approach to Review Legacy Schemas Based on Ontological Analysis

Raquel Lima Façanha¹ [0000-0003-4447-7057], Maria Cláudia Cavalcanti¹ [0000-0003-4965-9941],
and Maria Luiza Machado Campos² [0000-0002-7930-612X]

¹ Military Institute of Engineering, Rio de Janeiro, Brazil
rlfacanha@yahoo.com.br, yoko@ime.eb.br

² Federal University of Rio de Janeiro, Rio de Janeiro, Brazil
mluiza@ppgi.ufrj.br

Abstract: Usually, data schemas are the only documentation available for legacy data. Information Technology (IT) artifacts, such as conceptual schemas, if existent, are often outdated. This leads to inconsistencies and ambiguities, as well as difficulties in reusing data. This work proposes an approach for reviewing data schemas based on ontological analysis, which considers each concept according to its nature, capturing more precisely its essence and generally improving semantic richness and precision. The idea is to provide a systematic procedure to annotate legacy data, starting with its conceptual schema, and thus to contribute to generate more consistent conceptual modeling artifacts. In order to illustrate the proposed procedure, the Unified Foundational Ontology (UFO) is used as a theoretical reference for annotating a real data schema in the Legal domain.

Keywords: Ontological analysis, conceptual modeling, legacy data.

Towards a holistic documentation and wider use of Digital Cultural Heritage

Marinos Ioannides and Robert Davies¹

¹ Cyprus University of Technology, Limassol, Cyprus
marinos.ioannides@cut.ac.cy, robert.davies@cut.ac.cy

Abstract: This paper reviews work currently undertaken and planned to develop a more holistic approach to e-documentation of Cultural Heritage, thereby addressing the needs of a wider range of existing and potential audiences in the digital sphere. Building on the work of the ViMM Coordination and Support Action, funded under Horizon 2020, Digital Heritage Research Laboratory (DHRLab) at Cyprus University of Technology (CUT) has committed its research agenda for the years to come to the development these approaches, settings in train this vital process through three main mechanisms, aiming to create a holistic framework for DCH by carrying out the wide range of collaborative and multi-disciplinary research needed within an overall construct of advanced documentation:

1. The Europeana Task Force on Advanced Documentation of 3D digital assets
2. The UNESCO Chair on Digital Heritage
3. The Mnemosyne European Research Area Chair on Digital Heritage (Horizon 2020).

Keywords: Digital heritage, holistic, e-Documentation.

Graph Matching based Semantic Search Engine

Mamdouh Farouk¹, Mitsuru Ishizuka², and Danushka Bollegala³

¹Computer Science Department, Faculty of Computers and Information,
Assiut University, Assiut, Egypt
mamfarouk@aun.edu.eg

²National Institute of Informatics, Japan
ishizuka@nii.ac.jp

³Department of Computer Science, University of Liverpool, Liverpool, UK
danushka.bollegala@liverpool.ac.uk

Abstract: Explosive growth of the Web has made searching Web data a challenging task for information retrieval systems. Semantic search systems that go beyond the shallow keyword matching approaches and map words to their conceptual meaning representations offer better results to the users. On the other hand, a lot of representation formats have been specified to represent Web data into a semantic format. We propose a search engine for searching Web data represented in UNL (Universal Networking Language). UNL has numerous attractive features to support semantic search. One of the main features is that UNL does not depend on domain ontology. Our proposed search engine is based on semantic graph matching. It includes semantic expansion for graph nodes and relation matching based on relation meaning. The search results are ranked depending on the semantic similarity between the user query and the retrieved documents. We developed a prototype implementing the proposed semantic search engine, and our evaluations demonstrate its effectiveness across a wide-range of semantic search tasks.

Keywords: Search engine, semantic search, graph matching, UNL.

SKOS-based Concept Expansion for LOD-enabled Recommender Systems

Lisa Wenige¹, Geraldine Berger, and Johannes Ruhland

¹Chair of Business Information Systems
Friedrich-Schiller-University Jena, Germany
lisa.wenige@uni-jena.de

Abstract: This paper presents a concept expansion strategy for Linked Open Data-enabled recommender systems (LDRS). This strategy is based on annotations from Simple Knowledge Organization System (SKOS) vocabularies. To this date, the knowledge structures of SKOS graphs have

not yet been thoroughly explored for item similarity calculation in content-based recommender systems (RS). While some researchers have already performed an unweighted concept expansion on skos:broader links, the quantification of the relatedness of concepts from SKOS graphs with quality issues, such as the DBpedia category system, should be further investigated to improve recommendation results. For this purpose, we apply our approach in conjunction with a suitable concept-to-concept similarity metric and test it on three different LDRS datasets from the multimedia domain (i.e., movie, music and book RS). The results showed that our approach has a diversifying effect on result lists, while at least providing the same level of accuracy as a system running in non-expansion mode.

Keywords: SKOS, linked open data, recommender systems, and concept expansion.

*Navigating OWL 2 Ontologies through Graph Projection**

Ahmet Soylu^{1, 2} and Evgeny Kharlamov³

¹ Norwegian University of Science and Technology, Gjøvik, Norway

ahmet.soylu@ntnu.no

² SINTEF Digital, Oslo, Norway

³ University of Oxford, Oxford, the UK

evgeny.kharlamov@cs.ox.ac.uk

Abstract: Ontologies are powerful, yet often complex, assets for representing, exchanging, and reasoning over data. Particularly, OWL 2 ontologies have been key for constructing semantic knowledge graphs. Ability to navigate ontologies is essential for supporting various knowledge engineering tasks such as querying and domain exploration. To this end, in this short paper, we describe an approach for projecting the non-hierarchical topology of an OWL 2 ontology into a graph. The approach has been implemented in two tools, one for visual query formulation and one for faceted search, and evaluated under different use cases.

Keywords: OWL 2, ontologies, graph navigation, knowledge graphs.

* Funded by EU H2020 TheyBuyForYou (780247) and FP7 Optique (318338) projects.

Linked Data Live Exploration with Complete Results

Samita Bai¹[0000-0003-0024-3689], Sharaf Hussain¹[0000-0003-3901-9588],
and Shakeel Khoja¹[0000-0002-2275-7464]

¹ Institute of Business Administration (IBA)
Garden Road, Karachi-74400, Pakistan
{sbai, shussain, skhoja}@iba.edu.pk

Abstract: Linked Data is one of the emerging ways to publish and link structured and machine-processable data on the Web, however the existing techniques to perform live query Linked Data are based on recursive URI look-up process. These techniques contain a limitation for the query patterns having subject unbound and object containing a foreign URI. In such cases, the live query does not produce any answers to the query as the querying process could not be initiated due to unavailability of subject field in the triple pattern. In this paper, we make use of backlinking to extract and store foreign URIs and using this information for executing the queries live where the subject is un-bound.

Keywords: Linked Data, live querying, backlinking.

The Genesis of EngMeta: a metadata model for research data in Computational Engineering

Björn Schembera¹[0000-0003-2860-6621] and Dorothea Iglezakis²[0000-0002-8524-0569]

¹ High Performance Computing Center, Nobelstr. 19, 70569 Stuttgart, Germany
schembera@hlrs.de

² University Library Stuttgart, Holzgartenstr. 16, 70174 Stuttgart, Germany
dorothea.iglezakis@ub.uni-stuttgart.de

Abstract: In computational engineering, numerical simulations produce huge amounts of data. To keep this research data findable, accessible, inter-operable and reusable, a structured description of the data is indispensable. This paper outlines the genesis of EngMeta -- a metadata model designed to describe engineering simulation data with a focus on thermodynamics and aerodynamics. The metadata model, developed in close collaboration with engineers, is based on existing standards and adds discipline-specific information as the main contribution. Characteristics of the observed system offer researchers important search criteria. Information on the hardware and software used and the processing steps involved helps to understand and replicate the data. Such metadata are crucial to keeping the data FAIR and bridging the gap to a sustainable research data management in computational engineering.

Keywords: Research data management, metadata, big data, high performance computing, simulation, computational engineering.

Track on Digital Libraries, Information Retrieval, Big, Linked, Social and Open Data

Analysing and Visualising Open Data within the Data & Analytics Framework

Fallucchi Francesca^{1,2} [0000-0002-3288-044X], Michele Petito³ [0000-0002-5463-5322],
and De Luca Ernesto William^{1,2} [0000-0003-3621-4118]

¹ DIII, Guglielmo Marconi University, Rome, Italy

² DIFI, Georg Eckert Institute Braunschweig, Germany

³ Università di Pisa, Italy

[f.fallucchi, ew.deluca]@unimarconi.it

[fallucchi, deluca]@gei.de

michele.petito@unipi.it

Abstract: The principles of open data and the five-star model allow companies to develop low-cost services and Public Administrations (PA) to improve efficiency. However, the process of implementing open data models and principles is not easy unless it is supported by an appropriate technology platform. Today there is a huge number of technological platforms which each promise to be the ideal solution for opening data. Current solutions (commercial or free) do not provide users with easy access to data, nor tools for analysing and displaying data. In this paper, we discuss the potential of the DAF (Data Analytics Framework), a project based on big data, which was created by the Italian government in 2017 and which fosters the integration and standardisation of data, as well as providing three powerful tools for analysis and data visualisation. The paper will then illustrate a concrete case of dashboard development within the DAF, released at an important hackathon organised by the Italian PA sector in October 2017. The project serves as a use case in DAF implementation, where its analytical tools are used for data analysis & visualisation. They also translate a large amount of data into simple representations and use clear and effective language.

Keywords: Big Data, DAF, Data & Analytics Framework, data visualization, dashboard, Open Data, Linked Open Data.

Formalizing Enrichment Mechanisms for Bibliographic Ontologies in the Semantic Web

Helena Simões Patrício^{1,2} [0000-0002-4261-126X], Maria Inês Cordeiro¹ [0000-0002-6897-961X],
and Pedro Nogueira Ramos^{2,3} [0000-0003-0231-4356]

¹ Biblioteca Nacional de Portugal, Lisboa, Portugal

² Information Sciences, Technologies and Architecture Research Center (ISTAR-IUL)

³ University Institute of Lisbon (ISCTE-IUL), Portugal

Abstract: This paper presents an analysis of current limitations to the reuse of bibliographic data in the Semantic Web and a research proposal towards solutions to overcome them. The limitations identified derive from the insufficient convergence between existing bibliographic ontologies and the principles and techniques of linked open data (LOD); lack of a common conceptual framework for a diversity of standards often used together; reduced use of links to external vocabularies and absence of Semantic Web mechanisms to formalize relationships between vocabularies, as well as limitations of Semantic Web languages for the requirements of bibliographic data interoperability. A proposal is advanced to investigate the hypothesis of creating a reference model and specifying a superontology to overcome the misalignments found, as well as the use of SHACL (Shapes Constraint Language) to solve current limitations of RDF languages.

Keywords: Linked Open Data, bibliographic data, semantic web, SHACL, LOD validation, ontologies, reference model, bibliographic standards.

GLOBDEF: a framework for dynamic pipelines of semantic data enrichment tools

Maria Nisheva-Pavlova^[0000-0002-9917-9535] and Asen Alexandrov^[0000-0003-0619-8487]

Faculty of Mathematics and Informatics,
Sofia University St. "Kliment Ohridski", Sofia, Bulgaria
{marian, asenaleksandrov}@fmi.uni-sofia.bg

Abstract: Semantic data enrichment adds information to raw data to allow computational reasoning based on the meaning of data. With the introduction of Linked Data a lot of work is spent on combining existing tools for specific enhancement needs into multi-domain reusable enhancement pipeline. As part of the GloBIG project, we are working on the development of a framework for data enhancement, which attempts to be domain-agnostic and dynamically configurable. It works with pluggable enhancement modules, which are dynamically activated to create on-the-fly pipelines for data enhancement. Our research goal is to find a way for processing large amounts of data and automatically enhancing it while leveraging variety of domain knowledge sources and tools by selecting and using the most suitable ones according to the data. In this paper we present our proof-of-concept implementation of the so called GLOBDEF framework and discuss the challenges and next steps on its development.

Keywords: Linked Open Data, semantic annotation, semantic enhancement, metadata, ontology.

Ontologies for Data Science: on its application to data pipelines

Miguel-Ángel Sicilia, Elena García-Barriocanal, Salvador Sánchez-Alonso, Marçal Mora-Cantallops,
and Juan-José Cuadrado

Computer Science Department, University of Alcalá,
Polytechnic Building. Ctra. Barcelona km. 33.6,
28871 Alcalá de Henares (Madrid), Spain
{msicilia, elena.garciab, salvador.sanchez, marcal.mora, jjcg}@uah.es

Abstract: Ontologies are usually applied to drive intelligent applications and also as a resource for integrating or extracting information, as in the case of Natural Language Processing (NLP) tasks. Further, ontologies as the Gene Ontology (GO) are used as an artifact for very specific research aims. However, the value of ontologies for data analysis tasks may also go beyond these uses and span supporting the reuse and composition of data acquisition, integration and fusion code. This requires that both data and code artifacts support meta-descriptions using shared conceptualizations. In this paper, we discuss the different concerns in semantically describing *data pipelines* as a key reusable artifact that could be retrieved, compared and reused with a degree of automation if semantically consistent descriptions are provided. Concretely, we propose attaching semantic descriptions for data and analytic transformations to current backend-independent distributed processing frameworks as *Apache Beam*, as these already abstract out the specificity of supporting execution engines.

Keywords: Data science, ontologies, data pipelines, Apache Beam.

Relating Legal Entities via Open Information Extraction

Giovanni Siragusa¹, Rohan Nanda¹, Valeria De Paiva², and Luigi Di Caro¹

¹ University of Turin – Department of Computer Science,
Corso Svizzera 185, Turin, Italy
{siragusa, dicaro, nanda}@di.unito.it

² Nuance Communication Inc.,
Sunnyvale, California, U.S.A.
Valeria.dePaiva@nuance.com

Abstract: Concepts and relations within existing ontologies usually represent limited subjective and application-oriented views of a domain of interest. However, reusing resources and fine-grained conceptualizations is often challenging and requires significant manual efforts of adaptation to fit with unprecedented usages. In this paper, we present a system that makes use of recent Open Information Extraction technologies to unravel and explore corpus-centered unknown relations in the legal domain.

Keywords: Open Information Extraction, Natural Language Processing, ontologies, legal concepts, legal text, IATE.

*Ontology-based Information Retrieval: development of a semantic-based tool
for the media industry*

Ricardo Eito-Brun

Universidad Carlos III de Madrid, c/ Madrid, Getafe (Madrid), Spain
reito@bib.uc3m.es

Abstract: This paper describes the creation of an RDF ontology designed to support information retrieval needs of journalists and media professionals. The purpose of the ontology is to complete the automated extensions of query terms by using the relationships between the concepts and terms registered in the ontology. By using this ontology, end-users can identify additional concepts that are related to the selected topic, and incorporate new terms to the query that will be later launched against a full-text indexer based on SOLR. The ontology focuses on politics, and has been successfully tested with the collaboration of a large Spanish media company. The ontology contributes to a better recall of the search results.

Keywords: Human Computer Interaction, ontologies and controlled vocabularies, semantic ontologies in industry.

Track on Cultural Collections and Applications

Evaluating Data Quality in Europeana: metrics for multilinguality

Péter Király¹ [0000-0002-8749-4597], Juliane Stiller² [0000-0001-8184-6187], Valentine Charles³ [0000-0003-4085-5778],
Werner Bailer⁴ [0000-0003-2442-4900], and Nuno Freire⁵ [0000-0002-3632-8046]

¹ Gesellschaft für wissenschaftliche Datenverarbeitung mbH, Göttingen, DE
peter.kiraly@gwdg.de

² Humboldt-Universität zu Berlin, DE
juliane.stiller@ibi.hu-berlin.de

³ Europeana Foundation, The Hague, NL
valentine.charles@europeana.eu

⁴ Joanneum Research, Graz, AT
werner.bailer@joanneum.at

⁵ INESC-ID, Lisboa, PT
nuno.freire@tecnico.ulisboa.pt

Abstract: Europeana.eu aggregates metadata describing more than 50 million cultural heritage objects from libraries, museums, archives and audiovisual archives across Europe. The need for quality of metadata is particularly motivated by its impact on user experience, information retrieval and data re-use in other contexts. One of the key goals of Europeana is to enable users to retrieve cultural heritage resources irrespective of their origin and the material's metadata language. The presence of multilingual metadata description is therefore essential to successful cross-language retrieval. Quantitatively determining Europeana's crosslingual reach is a prerequisite for enhancing the quality of metadata in various languages. Capturing multilingual aspects of the data requires us to take into account the full lifecycle of data aggregation including data enhancement processes such as automatic data enrichment. The paper presents an approach for assessing multilinguality as part of data quality dimensions, namely completeness, consistency, conformity and accessibility. We describe the measures defined and implemented, and provide initial results and recommendations.

Keywords: Metadata quality, multilinguality, digital cultural heritage, Europeana, data quality dimensions.

The Benefits of Linking Metadata for Internal and External Users of an Audiovisual Archive

Victor de Boer^{1,2}, Tim de Bruyn^{1,2}, John Brooks^{1,2}, and Jesse de Vos²

¹ Vrije Universiteit Amsterdam, the Netherlands

² Netherlands Institute for Sound and Vision, Hilversum, the Netherlands

Abstract: Like other heritage institutions, audiovisual archives adopt structured vocabularies for their metadata management. With Semantic Web and Linked Data now becoming more and more stable and commonplace technologies, organizations are looking now at linking these vocabularies to external sources, for example those of Wikidata, DBPedia or GeoNames. However, the benefits of such endeavors to the organizations are generally underexplored. In this paper, we present an in-depth case study into the benefits of linking the “Common Thesaurus for Audiovisual Archives” (or GTAA) and the general-purpose dataset Wikidata. We do this by identifying various use cases for user groups that are both internal as well as external to the organization. We describe the use cases and various proofs-of-concept prototypes that address these use cases.

Keywords: Audiovisual metadata, Wikidata, case study.

Authify: the reconciliation of entities at scale

Philip E. Schreur¹ and Tiziana Possemato²

¹ Stanford University, Stanford, CA, USA

² @Cult, Rome, Italy – Casalini Libri, Fiesole, Italy

Abstract: Libraries’ shift to the semantic web has been underway for a number of years. Mellon funded projects such as Linked Data for Production (LD4P) [1] or the BIBFRAME European Workshop 2018 in Florence [2] show the commitment of national, public, and academic libraries, as well as vendors, to this transition. Libraries worldwide, however, are enmeshed in hundreds of millions of metadata records communicated through flat files (the MARC formats) [3]. The shift to linked data will require the conversion of these flat files to a semantically expressive model such as the Resource Description Framework (RDF) [4]. The conversion of such large amounts of semantically inexpressive data to semantically rich data will require automated enhancements in the conversion process. Data hidden within the flat files, such as role (author, illustrator, composer, etc.), can greatly aid with the reconciliation of entities within those files. Authify is one of the first tools available to libraries to both convert their metadata to linked data, but also enrich the reconciliation process with semantic data hidden within the MARC fields. As libraries look to convert their legacy data to linked data, Authify can help them move their data to the Web in as a semantically rich way as possible.

Keywords: Linked Data, reconciliation, MARC formats, Authify, libraries, semantic web.

Assessing the Preservation of Derivative Relationships in Mappings from FRBR to BIBFRAME

Sofia Zapounidou, Michalis Sfakakis, and Christos Papatheodorou

Department of Archives, Library Science and Museology, Ionian University, Corfu, Greece
{l12zapo, sfakakis, papatheodor}@ionio.gr

Abstract: Support of the exploration user task demands the explicit representation of bibliographic families and of content relationships. Seamless navigation through differently modelled bibliographic datasets presumes the existence of mappings. Semantic interoperability through mappings will be evaluated using a testbed. This paper starts with the fine-tuning of a testbed for mappings from FRBR to BIBFRAME. Two *Gold Standards datasets* have been created along with a mechanism for the mapping of core entities and the derivation relationship from FRBR to BIBFRAME. This first attempt has revealed that derivations expressed at the FRBR *Expression* level are mapped to BIBFRAME more adequately than those expressed at the FRBR *Work* level.

Keywords: BIBFRAME, derivative bibliographic relationships, FRBR, linked data, interoperability, mappings.

Metadata Standards for Palm Leaf Manuscripts in Asia

Nisachol Chamnongsri

School of Information Technology, Suranaree University of Technology, Thailand
nisachol@sut.ac.th

Abstract: The goal of this research is to facilitate, as effectively as possible, user access to and use of the knowledge recorded on palm leaf manuscripts (PLMs). At the same time, the schema should serve as a standard information structure to be used in the management of PLMs and other digitized ancient documents. This will also make the linking of Asian cultural heritage and wisdom with those of countries in this region possible via the internet. Accordingly, this research aims to develop metadata schema for the management of PLMs collections to increase efficiency in the search, access, use, and management. There are four parts in this study: (1) the current state of PLMs management in Asia and the use of PLMs metadata schema in working projects were investigated, then (2) the elements were analyzed and grouped by functions, (3) the core elements were matched to KKUPLMMs 2012, 2015 and IFLA LRM User Tasks and (4) a Focus group was set up to evaluate the framework.

Keywords: Metadata, palm leaf manuscripts, ancient documents, cultural heritage.

Track on Knowledge IT Artifacts (KITA) in Professional Communities and Aggregations (KITA 2018)

Knowledge Artifacts for the Health: the PERCIVAL project

Fabio Sartori, Riccardo Melen, Matteo Lombardi, and Davide Maggiotto

REDS Lab - Department of Computer Science, Systems and Communication
University of Milan – Bicocca
viale Sarca, 336 20126 – Milan, Italy
sartori@disco.unimib.it

Abstract: Quality of life (QoL) of patients affected by chronic diseases and their caregivers is a very important and inter-disciplinary research topic. From recent literature, it emerges the need for new methodologies capable to reduce the impact of a chronic disorders on everyday life of affected people and their relatives, especially when they are geographically far from care centers: the *PERsonal Care Instructor and VALuator* (PERCIVAL) project, collaboration between the REDS Lab and Educational Factory srl, is a first attempt to build up an integrated environment to promote the sharing, deliberation and monitoring of decisions about different aspects of chronic diseases among all the actors involved.

Keywords: Wearable devices, knowledge artifacts, e-Health, Quality of Life.

Artfacts: a platform for making sense of and telling stories with cultural objects

Leonardo de Araújo¹

Universität Bremen, Bibliothekstraße 1, 28359 Bremen, Germany
araujo@informatik.uni-bremen.de

Abstract: This paper presents the conceptualization, implementation, and evaluation of a Fast-speed IT Platform called *Artfacts*, which was designed within the context of the Two-speed IT Infrastructure, where a foundational, stable, and slow infrastructure is complemented by a creative, experimental, and agile additional infrastructure capable of promptly responding to the needs of communities. The platform is an attempt to digitally incorporate strategies for making sense and reusing digital collections and mitigate problems concerning specialized knowledge required for profiting from the affordances of data repositories as a creative material. In this sense, through the cartography of information, the platform aims at widening the participation of individuals with no technical background in the

development and maintenance process of interpretive applications, no matter whether within cultural institutions or events such as hackathons for cultural heritage. Artifacts intermediates the reinterpretation of cultural datasets and the fabrication of interpretive applications by means of a flexible, general, and interoperable data model that is able to adapt to the demands of storytellers, and an open-ended Object-Oriented UI that enables analysis and experimentation by arranging and rearranging data elements into digital narratives.

Keywords: Digital storytelling, data model, digital collections, heritage interpretation, hackathons, knowledge map, GLAMs.

A Semantic-based Metadata Schema to Handle System and Software Configuration

Ricardo Eito-Brun

Universidad Carlos III de Madrid, c/ Madrid, Getafe (Madrid), Spain
reito@bib.uc3m.es

Abstract: Configuration management is a key process in the system and software engineering. Product integrity is a key requirement in the development of software-based solutions, and Configuration Management defines the set of practices aimed to ensure the consistency and coherence of the product during its full life cycle. Although there are different standards that establish the principles of Configuration Management, in complex projects that require the interaction of several entities and companies, a more precise, detailed specification about how to report, compare and handle configuration data is needed. This constitutes an interesting opportunity for metadata management professionals. This contribution presents the development of a tool for managing configuration management data. The case study - developed in the context of a system engineering company – makes use of semantic web languages (RDF, OWL) and technologies to support engineers in the registration, analysis, reporting and auditing of products' configuration. The solution defines different metadata used to handle configuration status, and a technical solution to handle them.

Keywords: Configuration management, RDF, ontologies, information management.

Track on Digital Humanities and Digital Curation (DHC)

Connecting and Mapping LOD and CMDI through Knowledge Organization

Fallucchi Francesca^{1,2} [0000-0002-3288-044X] and De Luca Ernesto William^{1,2} [0000-0003-3621-4118]

¹ DIII, Guglielmo Marconi University, Rome, Italy

² DIFI, Georg Eckert Institute Braunschweig, Germany

[f.fallucchi, ew.deluca]@unimarconi.it

[fallucchi, deluca]@gei.de

Abstract: This paper explains the connection and mapping of knowledge representations between RDF and CMDI. Therefore, the challenge is to create a bridge between Linked Open Data (LOD) and the Component MetaData Infrastructure (CMDI) to ensure that the limits of the two paradigms are compensated and strengthened to create a new hybrid approach. While on the one hand, CMDI is easier to use for modelling purposes, the Metadata is not descriptive enough for a document to be easily discoverable using Linked Data (LD) technologies to publish and to enrich the document's content. Yet on the other hand, the explicit semantics and high interoperability of LOD have many advantages, but its modelling process is too complex for non-expert users. Here we show how knowledge organization plays a crucial role in this issue.

Keywords: Component MetaData Infrastructure (CMDI), Linked Open Data (LOD), metadata for language resources, digital humanities, Knowledge Organization (KO).

Creating CMDI-Profiles for Textbook Resources

Fallucchi Francesca^{1,2} [0000-0002-3288-044X], Hennicke Steffen² [0000-0001-8038-8081],
and De Luca Ernesto W.^{1,2} [0000-0003-3621-4118]

¹ DIII, Guglielmo Marconi University, Rome, Italy

² DIFI, Georg Eckert Institute Braunschweig, Germany

[f.fallucchi, ew.deluca]@unimarconi.it

[fallucchi, deluca, hennicke]@gei.de

Abstract: This paper analyses the establishment of a common infrastructure standard covering metadata, content, and inferred knowledge to allow collaborative work between researchers in the humanities. Interoperability between heterogeneous resources and services is the key for a properly functioning infrastructure. In this paper, we present a digital infrastructure of our textbook-related services and data, which are available and open for researchers worldwide. In this process we adhere

to established standards and provide APIs for other services. In order to integrate our resources and tools into the CLARIN infrastructure and make them discoverable in the VLO (Virtual Language Observatory), we decided to use CMDI (Component MetaData Infrastructure). We focus in this paper on the creation process for a CMDI metadata profile which fulfils the needs of our projects.

Keywords: Component MetaData Infrastructure (CMDI), VLO (Virtual Language Observatory), CLARIN, textbook, metadata for language resources, digital humanities.

Track on European and National Projects

*Towards a Knowledge Graph based Platform for Public Procurement**

Elena Simperl¹, Oscar Corcho², Marko Grobelnik³, Dumitru Roman⁴, Ahmet Soylu^{4**}, María Jesús Fernández Ruiz⁵, Stefano Gatti⁶, Chris Taggart⁷, Urška Skok Klima⁸, Annie Ferrari Uliana⁹, Ian Makgill¹⁰, and Till Christopher Lech⁴

¹ University of Southampton, Southampton, the UK

² Universidad Politécnica de Madrid, Madrid, Spain

³ Jožef Stefan Institute, Ljubljana, Slovenia

⁴ SINTEF Digital, Oslo, Norway

⁵ Ayuntamiento de Zaragoza, Zaragoza, Spain

⁶ Cerved Group Spa US, Milano, Italy

⁷ OpenCorporates Ltd, London, the UK

⁸ Ministrstvo za javno upravo, Ljubljana, Slovenia

⁹ OESIA Networks SL, Madrid, Spain

¹⁰ OpenOpps Ltd, London, the UK

Abstract: Procurement affects virtually all sectors and organizations particularly in times of slow economic recovery and enhanced transparency. Public spending alone will soon exceed EUR 2 trillion per annum in the EU. Therefore, there is a pressing need for better insight into, and management of government spending. In the absence of data and tools to analyse and oversee this complex process, too little consideration is given to the development of vibrant, competitive economies when buying decisions are made. To this end, in this short paper, we report our ongoing work for enabling procurement data value chains through a knowledge graph based platform with data management, analytics, and interaction.

Keywords: Procurement, knowledge graphs, analytics, interaction.

* This work is funded by EU H2020 TheyBuyForYou project (780247).

** Corresponding author. Email: ahmet.soylu@sintef.no

Metadata for Large-scale Research Instruments

Vasily Bunakov¹

¹ Science and Technology Facilities Council, Harwell OX11 0QX, United Kingdom

vasily.bunakov@stfc.ac.uk

Abstract: The work outlines diverse effort of a few initiatives for metadata and attribution mechanisms that can be used for large-scale instruments hosted by shared research facilities. Specifically, the role of persistent identifiers and associated metadata is considered, in relation to cases where the use of references to large-scale instruments can support research impact studies and Open Science agenda. A few routes for the adoption of large-scale instruments metadata are outlined, with indication of their advantages and limitations.

Keywords: large-scale instruments, research facilities, research attribution, persistent identifiers, research information management, impact studies, Open Science.

Track on Agriculture, Food and Environment

*Identification and Exchange of Regulated Information on Chemicals: from metadata to Core Vocabulary**

Alicja Agnieszka Dys^[0000-0001-8464-0221]

European Chemicals Agency
alicja.agnieszka.dys@gmail.com

Abstract: Regulatory bodies perform risk assessments of chemicals and produce regulatory outcomes: evaluations and decisions on chemicals and conditions of their use. Access to scientifically proven and already regulated information becomes crucial for their efficient work and consistent decisions. Exchanging and reusing information relies on common understanding of the main concepts. Here is the challenge: even if the regulations and industry standards provide definitions of chemical substance, the interpretation poses some issues. This paper introduces a concept of Regulated Substance and aims to highlight the complexity of implementing semantic interoperability on regulated information between different parties. The regulatory activities of European Chemicals Agency (ECHA) overlap, follow or trigger activities performed by other authorities. Capabilities to exchange the information and having access to shared databases can increase regulatory benefits. The common initiative of European Commission, Publication Office, and EU agencies is looking into possibilities to exchange the information. One of the tools promoted by Publication Office – Core Vocabularies – is meant to facilitate interoperability between authorities. The initiative will build foundations for access to the public repositories of the regulated information and non-confidential scientific data for academia and researchers via Open Linked Data.

Keywords: Chemical substance, European Chemicals Agency (ECHA), regulated information, semantic interoperability, Core Vocabulary.

* The above represents the opinion of the author and is not an official position of the European Chemicals Agency.

Semantics for Data in Agriculture: a community-based wish list

Caterina Caracciolo¹, Sophie Aubin², Brandon Whitehead³, and Panagiotis Zervas⁴

¹ Food and Agriculture Organization of the United Nations, Rome, Italy
caterina.caracciolo@fao.org

² INRA, UAR 1266 DIST Délégation Information Scientifique et Technique, Versailles, France
sophie.aubin@inra.fr

³ CABI, Wallingford, UK
b.whitehead@cabi.org

⁴ Agroknow, Athens, Greece
pzervas@agroknow.com

Abstract: The paper reports on activities carried within the Agrisemantics Working Group of the Research Data Alliance (RDA). The group investigated on what are the current problems research and practitioners experience in their work with semantic resources for agricultural data and elaborated the list of requirements that are the object of this paper. The main findings include the need to broaden the usability of tools so as to make them useful and available to the variety of profiles usually involved in working with semantics resources; the need to online platform to lift users from the burden of local installation; and the need for services that can be integrated in workflows. We further analyze requirements concerning the tools and services and provide details about the process followed to gather evidence from the community.

Keywords: Semantics, agricultural data, vocabularies, ontologies.

Development of Methodologies and Standardized Services for Supporting Forest Economies

Thomas Katagis¹, Nikolaos Grammalidis², Evangelos Maltezos³, Vasiliki Charalampopoulou³,
and Ioannis Z. Gitas¹

¹ Laboratory of Forest Management and Remote Sensing, AUTH, Thessaloniki, Greece

² Information Technologies Institute, CERTH, Thessaloniki, Greece

³ Geosystems Hellas S.A., Athens, Greece

thkatag@for.auth.gr

Abstract: In the Mediterranean region, many types of forests are non-productive or degraded, although they could substantially contribute to growth of local economies. In Greece, 30% of the total area is covered by forests, however their contribution to the GDP is almost non-existent. An example is the chestnut production in Thessaly region of Greece, and especially in Mouzaki municipality, which is almost abandoned due to insufficient agricultural policies concerning establishment of alternative crops, and consequently leads to loss of potential income for the rural economy. The ARTEMIS project, funded by the Greek Secretariat for Research and Technology, aims at delivering an innovative information platform providing systematically high quality Earth Observation based products and services for monitoring forest health and supporting eventually the growth of forestry related economy and market. The architecture of the proposed platform will incorporate new OGC / ISO technologies, while the applicability of existing metadata standards for management of geospatial datasets will be evaluated. A pilot implementation of the developed system will be conducted in a selected area in Thessaly region of Greece.

Keywords: Forest economics, forest monitoring, Earth Observation, forestry metadata.

Track on Open Repositories, Research Information Systems and Data Infrastructures

Open Citation Content Data

Mikhail Kogalovsky^c [0000-0003-3791-3791], Thomas Krichel^a, Victor Lyapunov^a, Oxana Medvedeva^a,
Sergey Parinov^{ab} [0000-0001-8333-2657], and Varvara Sergeeva^a

a) Russian Presidential Academy of National Economy and Public Administration (RANEPA), Moscow, Russia

b) Central Economy Institute of RAS (CEMI RAS), Moscow, Russia

c) Market Economy Institute of RAS (MEI RAS), Moscow, Russia

sparinov@gmail.com

Abstract: There are several projects in the research community to make the citation data extracted from research papers more re-usable. This paper presents results from the CyrCitEc project to create a publicly available source of open citation content data extracted from PDF papers available at a research information system. To reach this aim the project team has created four outputs: 1) an open source software to parse papers' metadata and full text PDFs; 2) an open service to process papers' PDFs to extract citation data; 3) a dataset of citation data, including citation contexts (currently mostly for papers in Cyrillic); and 4) a visualization tool that provides users insight into the citation data extraction process and gives some control over the citation data parsing quality.

Keywords: Open Data, citation content, CyrCitEc, RePEc, Socionet.

The Case for Ontologies in Expressing Decisions in Decentralized Energy Systems

Elena García-Barriocanal, Miguel-Ángel Sicilia, and Salvador Sánchez-Alonso

Computer Science Department, University of Alcalá,
Polytechnic Building. Ctra. Barcelona km. 33.6,
28871 Alcalá de Henares (Madrid), Spain
{elena.garciab, msicilia, Salvador.sanchez}@uah.es

Abstract: Advanced in technologies for the decentralization of applications have enabled micro-grid energy systems that do not rely on central control and optimization but are controlled by their owners. This may eventually enable consumers or intermediaries to specify concrete and diverse conditions on the supply that not only concern throughput, price and stability but also elements as provenance (e.g. that energy is produced from renewable sources) or locality among others. Blockchain technologies have emerged as a possible solution for the integration of the stream of events generated by smart

meters and networks, providing tamper-proof ledgers for offerings, transactions and traces. However, that requires languages for expressing conditions that might become complex and have to be executed locally. In this paper, we review the state of decentralization in micro-grids and its requirements, and discuss the role of ontologies as support for expressing constraints in those networks.

Keywords: Microgrids, transactive energy, blockchain, decentralization, ontologies.

Posters

Metadata Based Datasets Placement in Smart Grids

Asma Zgolli, Christine Collet, and Houssem Chihoub

Univ. Grenoble Alpes, CNRS, Grenoble INP, LIG, CS 40700 - 38058 Grenoble Cedex 9

Abstract: To achieve the scalability and performance of workloads in Smart grid ecosystems, users need to choose the location of data sets on different types of data systems. Our approach to placement is based on the metadata of datasets and workloads, but also on the metadata describing the characteristics of existing data systems. The choice of target data systems is made using a recommendation algorithm. This paper provides a general overview of our approach. It presents the metadata developed, the placement algorithm and discusses the placement criteria.

Keywords: Data Lake, Smart grid, data placement, metadata, recommendation, Big Data.

1 Datasets placement in smart grids ecosystems

The context of this work is the specification of a general architecture of a data management system for smart grids. The variety and large volumes of data, the diversity of systems and tools, and the velocity of treatments are the most important challenges in orchestrating a smart grid ecosystem [4,5]. In addition, various analytical applications and business needs are emerging today at the crossing and the exploitation of data at several levels (raw and refined, current and historical, counters' measurement data and social data, etc.).

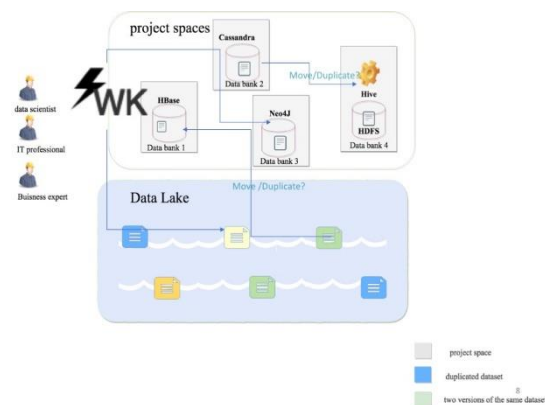


Figure 1 Smart grid Big data ecosystem

Modern smart grid big data ecosystem, relies on a data lake and several project spaces dedicated to user specific needs. The data lake designates the global storage space of information present within the ecosystem. This space is flexible enough to interact with the data, whether raw or highly refined. It allows to insert all the data, whatever their nature and origin because it does not impose a strict schema on incoming flows. It is based mainly on a distributed file system (such as Hadoop distributed file system HDFS [2] or Amazon S3[17]).

To give an operational dimension to the lake, a data bank is associated with a data bank management system that allows its processing within a so-called project space. In our context, we call a databank the databases and the data stores used to store datasets refined by business processes. Those datasets are

usually structured, personalized and qualified. Hence, a project space includes reliable and efficient workloads that execute SQL-like queries / more basic operations on datasets of the databank. In the project spaces, databanks can be private to one project or shared between teams. They are based on relational systems such as Teradata [20], document stores such as MongoDB [18], wide-column stores such as Cassandra [4], key/value stores such as Riak [16], and graph systems such as Neo4J [19]...). Databanks are also based on MapReduce based processing engines like apache Hive or on MPP based processing engines like apache Spark [14], or apache Drill [8]. However, these data bank management systems do not guarantee in the same way the performance, latency, scalability, consistency and availability. From the point of the data lake use, workloads of a project space need to access, join, and process data sets. Defining workloads needs an extensive knowledge about the data bank management systems. This can be an important issue given that not all users (e.g. IT specialists, data scientists, or non-IT professionals) have such skills, and the cost for developing applications/workloads would increase considerably. An issue in this context is the placement of the datasets, i.e. (i) in which data bank system to store the extracted data sets from the data lake in order to optimize their processing, and also (ii) in which data bank system to store the transformed data sets in order to optimize their future use.

2 Metadata based placement approach

In our approach, we propose to: (i) model the ecosystems data systems as system metadata, (ii) identify and collect metadata from user's datasets and workload and (iii) analyze this metadata to identify the suitable data systems for the storage and processing of datasets in order to assist the Smart grid ecosystems users.

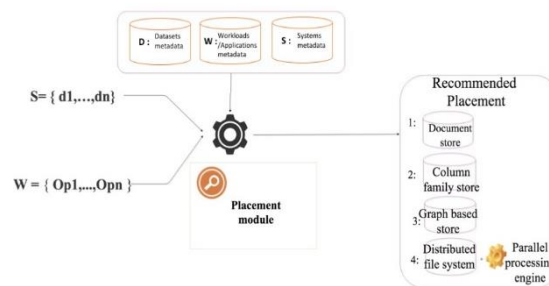


Figure 2 Recommendation for datasets placement

The set of target systems for the placement is obtained in three steps. The first step builds a placement solution space by inferring on the metadata. It takes as input a dataset $S = \{d_1, \dots, d_n\}$ and a workload W containing query operation $\{op_1, \dots, op_n\}$ on S . Based on the properties of S and the data operations done on S , the placement module (cf. figure 2) chooses elements of the solution space from the available data systems in the metadata base. The elements of the solution space should check the feasibility of the placement and its conformity with business environment & the data ecosystem.

The feasibility criteria allows identifying if the data store and the processing engine properties match with datasets and workload metadata. As for the conformity criteria, it determines if the placement result is valid in the business and data ecosystem. Indeed, given the complexity of the data ecosystem (#usages, #datasets versions, #level of processing, #access rights ...), we propose to consider semantic information in the placement logic. For this reason, we model that information as rules (placement rules) and we use them as a placement criterion. An example of a placement rule would be to restrain storing refined data in the lake. In our business context, it is important to preserve refined data in one

of the project spaces. The mechanism used to compute this solution space is based on inference on metadata represented as an RDF/RDFS graph using Logic rules (SWRL) and axioms.

In the second step, the placement module selects, from the solution space, effective placement candidates based on the performance of the workload evaluated using hybrid storage and processing cost models. In our approach, we are building a general cost model that combines existent cost models and supports different data systems. It allows us to compare the performance of heterogeneous big data systems (Relational dbms, NOSQL data stores, Parallel processing engines ...). The performance criterion measures response time of a workload execution on a giving dataset. Its value is estimated given cost parameters weighted to workload, data stores and datasets characteristics.

The final step consists on ranking and selecting the recommended placement systems to be returned to the user after comparing their different costs.

3 Metadata

The data placement module uses metadata as its data source. Indeed, in our data ecosystem, as well as in other information systems, metadata are varied and pervasive. we consider metadata represented in three abstract independent but complementary layers of the ecosystem referred to as Applications and Workloads, Data and Systems (cf. figure 4).

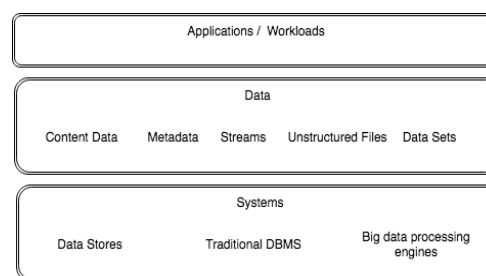


Figure 4 Metadata levels in smart grid big data ecosystem

Each layer describes a specific kind of metadata. Applications/workloads metadata characterizes query patterns, query models, query logical operators as well as query statistics. Other application metadata considered are semantic annotations and rules. The datasets metadata describes for instance (but not limited to): statistics, dataset size, values distribution, availability, location, schemas... Indeed, we characterize datasets with statistics about the number of records in a given dataset, the number of duplicated records or the number of records with null values and information about the quality of a dataset for instance: its level of processing or whether it has a duplicate stored in the ecosystem... As for systems metadata, it describes: the data models (e.g. column based data model, document based data model ...), the partitioning models (e.g. vertical partitioning, range based partitioning, consistent hashing), the distribution models, the query APIs and their logical operators, the physical storage and the physical operators...

4 Conclusion & related works

Data placement is a research subject that has been visited in a number of classic systems. Most of the existing applications [1,9,10,12,13,15] usually consider placement in a single database system/data store. Hybrid data placement approaches have been rarely discussed (we only found [3,7] that evaluates hybrid data placement), which emphasizes the novelty of our solution.

Data placement applications can be categorized into two groups depending on the data and the ecosystem of the placement. The first category [1,9,10,12,11,13] proposes approaches for placing chunks of data at the infrastructure level. The placement destination is usually a node server in a cluster. The second category [3,7,15] works on placement of dataset in ecosystems composed of autonomous data stores/ processing engine/ DBMS.

Our approach falls into the second category. We propose placement for datasets in a multi-store, multi-processing engines ecosystem based on different kinds of metadata. The main contributions of our research are: (i) Provide a unified view of metadata: description of data sets (sizes, distribution of values, availability, location, schemas, sizes, etc.), links between data sets, description of data systems and description of workloads...(ii) Collect and aggregate metadata on data sets (in multiple formats and from multiple sources) and workloads to facilitate subsequent use and (iii) Use of the modeled and extracted metadata in the recommendation for data placement algorithm.

ACKNOWLEDGMENTS

This work has been funded through the Industrial Research Chair Enedis (*) in Smart Grids. It has been carried out in collaboration with the Enedis Information Systems Division (Franck Atgie, Stephanie Delaunay, Sebastien Verdier Toureau).

REFERENCES

- [1] Rajashekhar M. Arasanal and Daanish U. Rumani. Improving Map-Reduce performance through complexity and performance based data placement in heterogeneous Hadoop clusters. In Chittaranjan Hota and Pradip K. Srimani, editors, Distributed Computing and Internet Technology, pages 115–125, Berlin, Heidelberg, 2013. Springer Berlin Heidelberg
- [2] Dhruva Borthakur. Hdfs architecture guide. Available at <https://hadoop.apache.org/docs/r1.2.1/hdfsdesign.html>, Accessed:2018–05–30, publisher: Apache Hadoop.
- [3] Francesca Bugiotti, Damian Burszty, Alin Deutsch, Ioana Ileana, and Ioana Manolescu. Invisible glue: scalable self-tuning multi-stores. In Conference on Innovative Data Systems Research(CIDR), 2015
- [4] Apache Cassandra Team. Apache cassandra, manage massive amounts of data, fast, without losing sleep. Available at <http://cassandra.apache.org/>, Accessed: 2018-05-28.
- [5] Housseem Chihoub and Christine Collet, "A Scalability Comparison Study of Data Management Approaches for Smart Metering Systems," 2016 45th International Conference on Parallel Processing (ICPP), Philadelphia, PA, 2016, pp. 474-483. doi: 10.1109/ICPP.2016.61[1]
- [6] Housseem Chihoub and Christine Collet, "iBig Hybrid Architecture for Energy IoT: When the power of Indexing meets Big Data Processing!," published in the 9th IEEE International Conference on Cloud Computing Technology and Science (CloudCom 2017), Hong Kong, dec 2017
- [7] Khoa Doan, Amidu Oloso, Kwo-Sen Kuo, T Clune, Hongfeng Yu, Brian Nelson, and Jian Zhang. Evaluating the impact of data placement to spark and scidb with an earth science use case. 12 2016
- [8] Apache Drill Team. "Apache Drill." Apache Drill - Schema-Free SQL for Hadoop, NoSQL and Cloud Storage, Apache, <http://drill.apache.org/>. Accessed:2018–06–08.
- [9] Mahdi Ebrahimi, Aravind Mohan, Andrey Kashlev, and Shiyong Lu. Bdap: A big data placement strategy for cloud-based scientific workflows. In Proceedings of the 2015 IEEE First International Conference on Big Data Computing Service and Applications, BIG-DATASERVICE '15, pages 105–114, Washington, DC, USA, 2015. IEEE Computer Society

- [10] Mohamed Y. Eltabakh, Yuanyuan Tian, Fatma Ozcan, Rainer Gemulla, Aljoscha Krettek, and John McPherson. Cohadoop: Flexible data placement and its exploitation in hadoop. Proc. VLDB Endow., 4(9):575–585, June 2011.
- [11] Tevfik Kosar and Miron Livny. A framework for reliable and efficient data placement in distributed computing systems. J. Parallel Distrib. Comput., 65(10):1146–1157, October 2005
- [12] K. Ashwin Kumar, Amol Deshpande, and Samir Khuller. Data placement and replica selection for improving co-location in distributed environments. CoRR, abs/1302.4168, 2013.
- [13] João Paiva, Pedro Ruivo, Paolo Romano, and Luís Rodrigues. AUTOPLACER: Scalable self-tuning data placement in distributed key-value stores. In Proceedings of the 10th International Conference on Autonomic Computing (ICAC 13), pages 119–131, San Jose, CA, 2013. USENIX.
- [14] Apache Spark Team. “Spark, Lightning-Fast Unified Analytics Engine.” Apache Spark™ - Unified Analytics Engine for Big Data, Apache, <http://spark.apache.org/>, Accessed:2018–06–08.
- [15] Abdul Quamar, K. Ashwin Kumar, and Amol Deshpande. Sword: Scalable workload-aware data placement for transactional workloads. In Proceedings of the 16th International Conference on Extending Database Technology, EDBT ’13, pages 430–441, New York, NY, USA, 2013. ACM.
- [16] Riak Team. Riak kv Available at <http://basho.com/products/riak-kv/>, Accessed: 2018-05-28.
- [17] Aws Team. Aws — amazon s3 — stockage de données en ligne dans le cloud. Available at <https://aws.amazon.com/fr/s3/>, Accessed: 2018-05-28
- [18] MongoDB.3.4 Team. Mongodb architecture guide. Available at <https://saipraveenblog.files.wordpress.com/2016/12/mongodbarchitectureguide.pdf>, Accessed:2017–11–28, June2017.
- [19] Neo4J Team. The neo4j graph platform – the 1 platform for connected data. Available at <https://neo4j.com/>, Accessed: 2018-05-28, Mai2018
- [20] Teradata Team. Teradata. Available at <https://www.teradata.fr/>, Accessed: 2018-05-28

EFSA Controlled Terminology – FoodEx2 Classification and Description System

D. Gibin, S. Ioannidou

Evidence Management Unit, EFSA
European Food Safety Authority

Via Carlo Magno 1A, 43126 Parma, Italy

Tel. +39 0521 036 111 | Fax +39 0521 036 110 | www.efsa.europa.eu

Abstract: The European Food Safety Authority (EFSA) aims to enhance the quality of its scientific outputs by giving direct access to scientific data and promoting the development of international collaborative platforms. In this context EFSA created and maintains more than 100 controlled terminologies. The use of controlled terminologies facilitates the aggregation of data during analysis and ensures interoperability between datasets. One of the most comprehensive controlled terminologies is the FoodEx2 – standardised food and feed classification and description system. FoodEx2 is used to transmit data to EFSA in relation to several food safety domains. This paper describes the FoodEx2 food classification and description system, the rationale for its development and its application in the remit of food and feed safety risk assessments. Finally, future perspectives on the use of FoodEx2 are discussed.

Keywords: Food classification, food description, controlled terminologies, XML, FoodEx2.

1 Introduction

Data from several data domains such as food consumption, chemical contaminants, pesticide residues, veterinary medicinal products residues (VMPP), zoonoses and food composition are submitted to EFSA from several data providers such as EU Member States competent authorities, the European Commission, industry, Consumers associations and Universities for use in EFSA's scientific assessments. EFSA collects and manages more than 40 million of data on an annual basis. To facilitate the aggregation of data during analysis EFSA uses controlled terminologies to ensure interoperability between datasets.

EFSA created and maintains more than 100 controlled terminologies, such as the 'PARAM' terminology which contains possible analytes (i.e. chemical contaminants, pesticides, veterinary medicinal products, additives, flavourings, microorganisms), the 'NUTS' terminology that includes the Nomenclature of Units for Territorial Statistics, the ANLYMD terminology regarding the analytical methods, etc. Each terminology contains a structured list of terms according to different classifications/authorities bodies (European Statistical Office (EUROSTAT), European and Mediterranean Plant Protection Organization (EPPO), etc.), legislative requirements and data collection needs. Controlled terminologies are language independent, meaning that items can be reported with the same code in different languages.

One of the most comprehensive controlled terminologies is the FoodEx2 – standardised food classification and description system (EFSA, 2015). The FoodEx2 aims at covering the need to describe food and feed in data collections across different food and feed safety domains. It ensures that all these data are submitted to EFSA in the same format and are therefore harmonised. FoodEx2 is used by EU Member States for submitting data on food consumption, chemical contaminants, pesticide residues, VMPP and food composition to EFSA.

2 FoodEx2 Food Classification and Description System

In an effort to introduce standardised food nomenclature across pan-European data collection activities, a Working Group (WG) including EFSA staff and external experts was established to develop a suitable food classification and description system with general applicability. The WG concluded that, while a comprehensive description system is available, no food classification was built with a scope broad enough to cover the needs of all domains relevant to intake or exposure assessments. Therefore, FoodEx2 has been developed and is described in the scientific document "Report on the development of a Food Classification and Description System for exposure assessment and guidance on its implementation and use" (EFSA, 2011a).

After its first release in 2011 (EFSA, 2011b), FoodEx2 was intensively tested with regard to the collection of food consumption and chemical occurrence data by several EU Member State organisations that send data to EFSA. In 2015 FoodEx2 was reviewed and revised in order to match the needs expressed by the users (EFSA, 2015). FoodEx2 is freely available for download and use¹.

FoodEx2 is EFSA's most comprehensive controlled terminology that follows a standard structure (EFSA, 2014). According to this structure terms follow a hierarchical parent- child relationship. A term can belong to one or more hierarchies and shall be linked to only a single parent term.

¹<https://www.efsa.europa.eu/en/data/data-standardisation>

The core FoodEx2 is the master hierarchy, which includes all the terms of the controlled terminology. Additional hierarchies that are defined as “reporting/analysis hierarchies” include a subset of the terms available in the master hierarchy, or can even change the parent-child relationship of the terms available in the master hierarchy. The available hierarchies currently in FoodEx2 are the Reporting, Exposure, Feed, Zoonoses, VetDrugRes, Botanicals and FeedAddExpo hierarchy. The Exposure hierarchy is the most commonly used for the collection of food consumption, chemical contaminants and pesticides residues on food.

The Exposure hierarchy of FoodEx2 consists of descriptions of a large number of individual food items aggregated into 21 food groups such as grain and grain-based products, fruit and fruit products, meat and meat products etc. Central to the system is a core list of food items or generic food descriptions that, in the majority of cases, represent the minimum level of detail. The description of individual foods can be complemented through the use of facets. Facets are used to add further detail, in relation to different aspects of foods, to the core terms. For example, facets can provide information about the processing a food has been subjected to, its packaging material and packaging format, its production method etc.

The FoodEx2 classification system enables occurrence data to be combined with food consumption data in order to allow the calculation of dietary exposure, an important step of the risk assessment work carried out in EFSA.

FoodEx2 is embedded in the standard data models (EFSA, 2013, EFSA, 2014) used by European Union (EU) Member State organisations, food industry, academia, etc. to transmit data to EFSA in relation to several food safety domains, e.g. for analytical results.

3 Challenges

FoodEx2 is an evolving system. In order to have it aligned with the market update as well as new regulations in the area of food and feed safety, maintenance needs to be constantly performed. This is achieved using versioning of the FoodEx2 terminology.

4 Future Perspectives

FoodEx2 is currently used at international level. Institutions like Food and Agriculture Organisation of the United Nations (FAO) and World Health Organisation (WHO) are using FoodEx2 to facilitate the collection of food consumption and food composition data on a global level to populate the Global Individual Food consumption data Tool (GIFT) database². In addition, other organisations in the USA, Canada, Australia, etc. are using FoodEx2 for the codification of data collected in different domains.

EFSA is participating in the new GODAN Working Group on “Publication and Alignment of Authoritative Vocabularies for Food”³ and is contributing with the FoodEx2 system. This WG is bringing together the key domain authorities and other partners who can provide the platforms and methodologies to work on and publish vocabularies. Its final objective is to provide a set of semantically interlinked URIs for key food product concepts as an infrastructural component that will facilitate the development of both software for actors in the food value chain and added-value integrated services that need to track food products. This activity will facilitate the use of FoodEx2 in software and tool development.

² <http://www.fao.org/gift-individual-food-consumption/en/>

³ <http://www.godan.info/working-groups/publication-and-alignment-authoritative-vocabularies-food-working-group>

Due to the international interest FoodEx2 is attracting, the translation of its catalogue in different languages could be a possible development, in the remit of the improvement of data sharing and openness. In addition, further to the already performed training/webinars⁴, future webinars could be an asset to enlarge the audience and increase the usability of FoodEx2.

5 Conclusions

EFSA is continuously working towards a more harmonised way of collecting data with the view to facilitate data sharing and exchange. In this remit, EFSA has developed controlled terminologies for allowing the comparison of data among different domains and the implementation of more detailed data analysis. FoodEx2, the most comprehensive controlled terminology of EFSA, is currently used by EU Member States organisation to transmit data to EFSA as well as at international level for the codification of data collected in different domains. EFSA is engaged in improving and promoting best practice regarding its use.

REFERENCES

- [1] European Food Safety Authority (EFSA), 2011a; Report on the development of a food classification and description system for exposure assessment and guidance on its implementation and use. EFSA Journal 011; 9 (12):2489. [84 pp.] doi:10.2903/j.efsa.2011.2489.
- [2] EFSA, 2011b; The food classification and description system FoodEx2 (draft revision 1). Supporting Publications 2011:215. [438 pp.].
- [3] EFSA, 2015; The food classification and description system FoodEx2 (revision 2). EFSA supporting publication 2015: EN-804. 90pp.].
- [4] EFSA, 2014; Guidance on Data Exchange version 2.0. EFSA Journals 2014; 12(12): 3945, 173 pp., doi: 10.2903/j.efsa.2014.3945
- [5] EFSA, 2013; Standard Sample Description ver. 2.0. EFSA Journal 2013; 11(10):3424, 114 pp., doi: 10.2903/j.efsa.2013.3424.

ABBREVIATIONS

VMPPR: veterinary medicinal products residues
EUROSTAT: European Statistical Office
EPPO: Mediterranean Plant Protection Organization
WG: Working Group
FAO: Food and Agriculture Organisation of the United Nations
WHO: World Health Organisation
GIFT: Global Individual Food consumption data Tool
GODAN: Global Open Data for Agriculture and Nutrition

⁴<https://www.efsa.europa.eu/en/events/event/180926>

Set Up of a Metadata Strategy – First step: indexing your data

Jean Miguet & Stephane Auger

France, Danone Research

jean.miguet@gmail.com, stephane.auger@danone.com

Abstract: Over the past two decades with the emergence of the big data deluge, the associated technologies and the ensuing opportunities, enterprises have realized that data are one of their key assets and, as such, needs to be managed by an adapted governance. A fundamental pillar within an effective data governance strategy is a clear and consistent metadata strategy. A metadata strategy impacts all aspects of the enterprise data lifecycle, and demands a strong implication of all involved actors, from both the technical and business sides. The first objective of our metadata strategy in Danone Research is ensuring a global understanding of data ecosystem we are working on. For this, we've implemented the following blocks to define and maintain the strategy: State of the Art, User Needs, Tools & Architecture, Ways of Working and Governance and Assessment & Monitoring. And more concretely, as our first tangible deliverable we chose to produce a (meta)Data Catalog. This catalog serves as a central inventory of all our data sources (both internal and external). Each data source has a variety of metadata tags that describe the context of the data, such as the source, where the data is located, the data owner, its quality, accessibility, regulatory constraints etc. All these metadata elements provide a wholistic picture of our data. And most importantly, it allows us to link our company's strategic objectives, priorities, and projects, with the data that are needed to meet these objectives. It allows us to quickly identify what data we have, what data we don't have and need to get, what data we can rely on to make good business decisions.

Keywords: Metadata strategy, data catalog, standards.

Author Index

A

Alexandrov, Asen	35
Aubin, Sophie	47
Auger, Stephane	64
B		
Bai, Samita	32
Bailer, Werner	38
Bārzdiņš, Juris	29
Berger, Geraldine	31
Bojārs, Uldis	29
Bollegala, Danushka	31
Brooks, John	38
Bunakov, Vasily	45
C		
Campos, Maria Luiza Machado	29
Caracciolo, Caterina	47
Cavalcanti, Maria Cláudia	29
Čerāns, Kārlis	29
Chamnongsri, Nisachol	40
Charalampopoulou, Vasiliki	48
Charles, Valentine	38
Chihoub, Houssein	52
Collet, Christine	52
Corcho, Oscar	45
Cordeiro, Maria Inês	34
Cuadrado, Juan-José	36
D		
Damigos, Matthew	27
Davies, Robert	30
De Araújo, Leonardo	41
de Boer, Victor	38
de Bruyn, Tim	38
De Luca, Ernesto William	34, 43
De Paiva, Valeria	36
de Vos, Jesse	38
Di Caro, Luigi	36
Dys, Alicja Agnieszka	47
E		
Eito-Brun, Ricardo	37, 42

España-Bonet, Cristina	28
F		
Façanha, Raquel Lima	29
Fallucchi, Francesca	34, 43
Farouk, Mamdouh	31
Freire, Nuno	38
G		
García-Barriocanal, Elena	36, 49
Gatti, Stefano	45
Gergatsoulis, Manolis	27
Gibin, Davide	60
Gitas, Ioannis Z.	48
Grądzki, Przemysław	28
Grammalidis, Nikolaos	48
Grasmanis, Mikus	29
Grobelnik, Marko	45
H		
Hussain, Sharaf	32
I		
Iglezakis, Dorothea	33
Ioannides, Marinos	30
Ioannidou, Sofia	60
Ishizuka, Mitsuru	31
K		
Kalogeros, Eleftherios	27
Katagis, Thomas	48
Kharlamov, Evgeny	32
Khoja, Shakeel	32
Király, Péter	38
Klima, Urška Skok	45
Kogalovsky, Mikhail	49
Krichel, Thomas	49
Kuziński, Dominik	28
L		
Lāce, Lelde	29
Lech, Till Christopher	45
Ledl, Andreas	27
Lombardi, Matteo	41
Lyapunov, Victor	49
M		

Maggiotto, Davide	45
Makgill, Ian	45
Maltezos, Evangelos	48
Medvedeva, Oxana	49
Melen, Riccardo	41
Miguet, Jean	64
Mora-Cantallops, Marçal	36
N		
Nanda, Rohan	36
Nisheva-Pavlova, Maria	35
O		
Ovčihņkova, Jūlija	29
P		
Papatheodorou, Christos	24, 40
Parinov, Sergey	49
Patrício, Helena Simões	34
Petito, Michele	34
Petras, Vivien	28
Possemato, Tiziana	39
R		
Ramos, Pedro Nogueira	34
Ramthun, Roland	28
Roman, Dumitru	45
Ruhland, Johannes	31
Ruíz, María Jesús Fernández	45
S		
Sánchez-Alonso, Salvador	36, 49
Sartori, Fabio	41
Schembera, Björn	33
Schreur, Philip E.	39
Sergeeva, Varvara	49
Sfakakis, Michalis	40
Sicilia, Miguel-Ángel	36, 49
Simperl, Elena	45
Siragusa, Giovanni	36
Sopek, Mirek	28
Šostaks, Agris	29
Soylu, Ahmet	32, 45
Sproģis, Artūrs	29
Steffen, Henniscke	43
Stiller, Juliane	28, 38

T		
Taggart, Chris	45
Trójczak, Rafał	28
Trypuz, Robert	28
U		
Uliana, Annie Ferrari	45
V		
van Genabith, Josef	28
W		
Waeber, Jonas	27
Wenige, Lisa	31
Whitehead, Brandon	47
Z		
Zapounidou, Sofia	40
Zervas, Panagiotis	47
Zgolli, Asma	52